PMIx: Process Management for Exascale Environments
What is PMIx?

PMIx v1.2
OMPI Spectrum OSHMEM
PGAS others
Exascale launch in < 30s

PMIx v2.x
OMPI Spectrum OSHMEM SOS PGAS others
Exascale launch in < 10s Orchestration

years go by…

MPICH
wireup support
dynamic spawn
keyval publish/lookup

PMI-1 PMI-2
PMIx v1.0
RM
SLURM ALPS
Exascale systems on horizon
Launch times long
New paradigms

2015 2016 2017
RM
SLURM JSM others
Exascale launch in < 10s
Orchestration
Three Distinct Entities

• **PMIx Standard**
  - Defined set of APIs, attribute strings
  - Nothing about implementation

• **PMIx Reference Library**
  - A full-featured implementation of the Standard
  - Intended to ease adoption

• **PMIx Reference Server**
  - Full-featured “shim” to a non-PMIx RM
  - Provides development environment

*Standards Doc under development!*
Where Are We?

• Launch scaling
  ▪ Wireup enhancements complete
  ▪ Fabric “instant on” enablement underway

• Results
  ▪ Tracks spawn propagation time
  ▪ Exascale in < 5 seconds
  ▪ 3rd party confirmation

[Extended version to be published]
Stage 1

*RM daemon, mpirun-daemon, etc.
PMIx Launch Sequence

*RM daemon, mpirun-daemon, etc.
Current Support (I)

- Typical startup operations
  - Put, get, commit, barrier, spawn, [dis]connect, publish/lookup
- Tool connections
  - Debugger, job submission, query
- Generalized query support
  - Job status, layout, system data, resource availability
- Event notification
  - App, system generated
  - Subscribe, chained
  - Preemption, failures, timeout warning, ...
- Logging
  - Status reports, error output
- Flexible allocations
  - Release resources, request resources
Current Support (II)

- **Network support**
  - Security keys, pre-spawn local driver setup

- **Obsolescence protection**
  - Automatic cross-version compatibility
  - Container support

- **Job control**
  - Pause, kill, signal, heartbeat, resilience support (C/R coordination)

- **Async definition of process groups**
  - Rolling startup/teardown
In Pipeline

- **Network support**
  - Fabric topology and status, traffic reports, fabric manager interaction

- **MPI Sessions support**
  - Rolling startup/teardown

- **Generalized data store**
  - Distributed key-value storage

- **Security**
  - Obtain and validate credentials for application/SMS

- **Power directives**

- **File system support**
  - Dependency detection
  - Tiered storage caching strategies

- **Debugger/tool support**
  - Automatic rendezvous
  - Single interface to all launchers
  - Co-launch daemons
  - Access fabric info, etc.

- **Cross-library interoperation**
  - OpenMP/MPI coordination

*Legend:*
- Red: Nearing completion
- Blue: Ramping up
- Green: Idling
Further Info

- PMIx BOF
  - Wed 12:15, Rms 210-212
- Enhanced SLURM launch scaling
  - Right here, Thurs 2pm
- Community sites
  - https://github.com/pmix